

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 12, 2005, 16:41:49 ; Search time 174 Seconds

(without alignments)  
915.268 Million cell updates/sec

Title: US-10-664-506a-5

Perfect score: 1607

Sequence: 1 MGWAMMLPULLIGISGL.....FFFAVVAQVKEKLNVEEG 311

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: UniProt\_sprot:\*  
2: UniProt\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1017	63.3	223	2	Q8CHS7 mus musculus
2	651.5	40.5	310	2	Q6UX59 homo sapien
3	651.5	40.5	325	2	Q9BTP9
4	650.5	40.5	325	2	Q6TANO
5	640.5	39.9	325	2	Q9UFM6
6	625.5	38.9	323	2	Q99J47
7	557	34.7	291	2	Q9Y3A1
8	466	29.0	326	2	Q9Y140
9	458	28.5	317	2	Q70732
10	331	20.6	311	2	Q65Z13
11	322	20.0	264	2	Q81M93
12	321	20.0	264	2	Q635B8
13	320	19.9	264	2	Q731G1
14	311.5	19.4	537	2	Q22787
15	309	19.2	264	2	Q6HE26
16	302.5	18.8	268	2	Q6H8P2
17	299.5	18.6	278	2	Q9HX59
18	295.5	18.4	339	1	DHS7_HUMAN
19	295.5	18.4	375	2	Q659E8
20	290.5	18.1	267	2	Q8EJMS
21	288.5	18.0	324	2	Q6I7R1
22	286.5	17.8	324	2	Q6I7R1
23	282	17.5	338	1	DHS7_MOUSE
24	277	17.2	273	2	Q6W1F0
25	276	17.2	336	2	Q6GP08
26	274	17.1	277	2	Q8YX01
27	271	16.9	268	2	Q6CNU5
28	265.5	16.5	263	2	Q9ZAB9
29	265.5	16.5	266	2	Q97DY5
30	263	16.4	298	2	Q87XV8
31	262.5	16.3	260	2	Q65HP3

32	260.5	16.2	263	2	Q71Y45	Q71Y45 listeria mo
33	258	16.1	312	2	Q9N126	Q9N126 bos taurus
34	258	16.1	316	2	Q8T197	Q8T197 dictyosteli
35	256	15.9	280	2	Q65F58	Q65F58 bacillus li
36	255.5	15.9	248	2	Q9RH22	Q9RH22 zymomonas m
37	255.5	15.9	263	2	Q8Y5S9	Q8Y5S9 listeria mo
38	254	15.8	303	2	Q6F7B8	Q6F7B8 actinobact
39	253	15.7	259	1	Y00Q_BACSU	Y00Q_BACSU
40	253	15.7	295	2	Q8RR58	Q8RR58 actinobact
41	252	15.7	295	2	P94129	P94129 actinobact
42	250.5	15.6	271	2	Q7UPD8	Q7UPD8 rhodospirell
43	249.5	15.5	311	2	Q9NYR8	Q9NYR8 homo sapien
44	249	15.5	259	2	Q9KCR3	Q9KCR3 bacillus ha
45	247	15.4	276	2	Q9ALU7	Q9ALU7 enterobacte

## ALIGNMENTS

RESULT 1

Q8CHS7 PRELIMINARY; PRT; 223 AA.

AC Q8CHS7/

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Similar to CGI-86 protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; NCBI\_TaxID=10090;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Mammary gland;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haefl F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stadlein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Topolycki S., Carninci P., Prange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Boshak S.A., McSwan P.U., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hultky S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Kravinsky M.I., Skalski U., Small D.E., Scherch A., Schein J.E., Jones S.J., Marra M.A.;

RA "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Mammary gland;

RA Strausberg R.;

CC Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

CC -1-SIMILARITY: Belongs to the short-chain dehydrogenases/reductases (SDR) family.

CC EMBL: BC039565; AAH39565.1; "

CC GO: GO:0016491; F:oxidoreductase activity; IEA.

CC GO: GO:0008152; P:metabolism; IEA.

DR InterPro: IPR002198; ADH\_short.

DR InterPro: IPR002347; ADH\_short\_C2.

DR Pfam: PF00106; adh\_short; 1.

DR PRINTS: PR00081; GDHRDH.

DR PRINTS: PR00080; SDRFAMILY.

DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.

DR Oxidoreductase.

KM SEQUENCE 223 AA; 2508 MW; 368381EADFA15B90 CRC64;

[illegible]

## PRELIMINARY; PRT; 325 AA

AC O9BTP3; 17, Created)  
AD 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, last annotation update)  
DS DKFZP660084 protein.  
GN Name=DKFZP660084;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Theria; Primates; Catarrhini; Homnidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschuler S.F., Ziegler B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Pamer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Tshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fehey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmitt J., Myers R.M., Butterfield Y.S.,  
RA Krzywniuk M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
LN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Lung;  
RC Strausberg R.;  
RN Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
[3]

RC TISSUE=Lung;  
RA Strassberg R.1  
RC Submitted (JUN-2001) to the EMBL/GenBank/DBD databases.  
LL -1- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases  
CC (SDS) family.  
DR EMBL: BC004126; AA04126.1; -  
DR EMBL: BC003679; AA03679.1; -  
DR HSSP: Q92FY9; 1F0H.

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2005, 16:50:09 ; Search time 41 Seconds  
(without alignments)  
729.839 Million cell updates/sec

Title: US-10-664-506a-5

Perfect score: 1607

Sequence: 1 MGWAMLMPLLLIGISGL.....FFPAVACGVKELNPEEG 311

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5.

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	640.5	39.9	2 T17307	hypothetical prote
2	311.5	19.4	2 T34380	hypothetical prote
3	307	19.1	2 T34378	hypothetical prote
4	299.5	18.6	2 E83152	probable short-cha
5	274	17.1	2 A61983	hypothetical prote
6	265.5	16.5	2 A11692	oxidoreductase hom
7	265.5	15.9	2 H97309	short-chain alcoh
8	255.5	15.9	2 A11321	oxidoreductase hom
9	253	15.7	2 A69965	ketoacyl reductase
10	249	15.5	2 E83858	oxidoreductase Bhl
11	245	15.2	2 AD2176	oxidoreductase all
12	242	15.1	2 E86788	oxidoreductase yme
13	241.5	15.0	2 F75374	probable ketoacyl
14	241	15.0	2 F90254	hypothetical prote
15	241	15.0	2 A95223	hypothetical prote
16	238.5	14.8	2 C55210	hectn protein - Ana
17	235.5	14.7	2 A62475	ketoacyl reductase
18	234.5	14.6	2 E70082	glucose 1-dehydrog
19	234	14.6	2 A28788	actinorhodin polyk
20	233	14.5	2 S17652	follicular lymphom
21	230.5	14.3	2 A82604	oxidoreductase Xf2
22	230	14.3	2 A81866	hypothetical prote
23	227	14.1	2 E82247	oxidoreductase, sh
24	225.5	14.0	2 H72219	3-oxoacyl-(acyl ca
25	225	14.0	2 C98087	oxidoreductase (EC
26	225	14.0	2 T24832	hypothetical prote
27	224.5	14.0	2 H87085	short chain alcoh
28	224	13.9	2 D70761	probable fatty-acy
29	223	13.9	2 E95283	probable [imported

30	222	13.8	2 H86809	oxidoreductase ypa
31	222	13.8	2 G84171	oxidoreductase [im
32	221.5	13.8	1 D69930	probable 3-oxoacyl
33	221.5	13.8	2 T24540	hypothetical prote
34	221	13.6	2 H70890	hypothetical prote
35	219	13.6	2 D97493	3-oxoacyl-(acyl ca
36	219	13.6	2 A62711	3-oxoacyl-(acyl ca
37	218.5	13.6	2 I40869	hypothetical prote
38	216.5	13.5	2 A9	ribitol 2-dehydrog
39	216.5	13.5	2 S07134	hypothetical prote
40	216.5	13.5	2 C95078	conserved hypother
41	215	13.4	2 H97945	oxidoreductase of
42	214.5	13.3	2 T45275	hypothetical prote
43	214	13.3	2 D70853	short-chain type d
44	213.5	13.3	2 B86634	oxidoreductase hom
45	212.5	13.2	2 A11128	internalin B - lfs

# ALIGNMENTS

RESULT 1  
T17307  
hypothetical protein DKFZp5660084.1 - human  
C/Species: Homo sapiens (man)  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
R/Bioecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, September 1999  
A/Reference number: Z16724  
A/Accession: T17307  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-325 <BLO>  
A/Cross-references: UNIPROT:Q9UFM6; EMBL:AL117567  
A/Experimental source: fetal kidney; clone DKFZp5660084  
C/Genetics:  
A/Note: DKFZp5660084.1

Query Match 39.9%; Score 640.5; DB 2; Length 325;  
Best Local Similarity 44.2%; Pred. No. 1.5e-45;  
Matches 136; Conservative 58; Mismatches 101; Indels 13; Gaps 5;

QY	3	WMAMLMPLLL-LGISGLFIYQEVSRMSKSAVONKVVITTDALISGLKECARVHTG	60
DB	19	ITSTALPLPLPGCLGVFGLFRLLQWVR---GKAYLRNAVAVVIGATSGKEGAKVFYAA	75
QY	61	GARLVICGKMERLENTYDAL-ISVADPSKTFPKLVLDLSDISCVDAKXVLDQYGC	119
DB	76	GARLVICGKRGALBELIRELTASHATKVTQTHRPVLTVPDLTDSGAIVAAABETLOCFGY	135
QY	120	VDLLINNASVYKVGPAHKISLELDKKIMDANYEGPTLTFTALPLNMISRRTOIVLVNTI	179
DB	136	VDLLINNASVYKVGPTIMDTVDVDAKMEINNYGPAITALLPSMIKRQGHVVAISSI	195
QY	180	QKGFPIPTTVAASGHAALGFDCIAPAEVEYDVIVISVSPFIR---SYHYVPEQGNW	236
DB	196	QKGMISIPFSAYAPKATQAFDCIAPAEVEYDVIVISVSPFIR---SYHYVPEQGNW	253
QY	237	EASIKMFPFRKLTGVGHVPEVAEENRTVRRKQVEFMNPIKAVVYVTFPPEPFAV	296
DB	254	--SRGVMDDTTTAQGSPEVADVIAAAGKKKKDDVTLADLPSLAVYLTALPAGLFFSL	311
QY	297	VACGVKEX 304	
DB	312	MASRAKE 319	

RESULT 2  
T34380  
hypothetical protein T25G12.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C:\Accession: T34380  
R:\Du\_Z.  
submitted to the EMBL Data Library, December 1995  
A>Description: The sequence of C. elegans cosmid T25G12.  
A.Reference number: 221515  
A.Accession: T34380  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-537 <DUZ>  
A.Cross-references: UNIPROT:Q22787, EMBL:U43283, PIRN:AAC69018.1, GSPDB:GN00028, CESP:T25G12  
C.Genetics:  
A.Experimental source: strain Bristol N2; clone T25G12  
A.Gene: CESP:T25G12.2  
A.Map position: X  
A.Introns: 59/2; 89/1; 136/2; 279/3; 342/2; 426/3; 475/2; 526/2

Query Match            19.4%; Score 311.5; DB 2; Length 537;  
Best Local Similarity 30.0%; Pred. No. 5,7e-18;  
Matches    88; Conservative    57; Mismatches    99; Indels    49; Gaps    9;

Dy         31 SKSAVQKRVVITDAISGLGKECARVFHTGGARLVLCGRMERLENLYDALISVADPSKT 90  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         72 SKLVNKXIVITGTASSGSGKLAFELYKRGAQVILLARSTEKLEI-----CAEILTKT 125  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Dy         91 F-----TPKVLDLSDISCVP--DVAKEDVCSCGVILNNMSVKYKPAHHSIELD 143  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         126 FLPNKKRTYYFPDITNDPKAPWAQIPK-----VVLLNNAGMSNRSGCDTTAAIH 177  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Dy         144 KKIMDANFEPITTLKALLPMMISRTOQLVLNNIOGKFGIPFTTYAASKHALGFED 203  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         178 RKAETNLFGHVQVIGSLISLK--SPDCIVTSSIQKVAIPRGVSASAKHALOGYFD 235  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Dy         204 CLRAEVEEDVAVISTSPFFIRSYH-----VYPEGNGWASIWKEFFPKLTYG 252  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         226 CLRAEHKMLHLIV--VSAGYINTGFSRALDTDGKVGVEBENGOK-----GY 281  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Dy         253 HPVEVAEEVMATVRKKQEVPMANPIPKAAAYVRFPEEF--FAVAVAGVYEK 304  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         282 SPEHARMISDAIRDVSDFDMAFPGARFALFLRYFWPTLLNTVALYTRGTQDQ 334  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 3  
T34378  
hypothetical protein T25G12.7 - Caenorhabditis elegans  
C:\Species: Caenorhabditis elegans  
C:\Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:\Accession: T34378  
R:\Du\_Z.  
submitted to the EMBL Data Library, December 1995  
A>Description: The sequence of C. elegans cosmid T25G12.  
A.Reference number: 221515  
A.Accession: T34378  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-562 <DUZ>  
A.Cross-references: UNIPROT:Q22785, EMBL:U43283, PIRN:AAC69022.1, GSPDB:GN00028, CESP:T25G12  
C.Genetics:  
A.Experimental source: strain Bristol N2; clone T25G12  
A.Gene: CESP:T25G12.7  
A.Map position: X  
A.Introns: 28/2; 58/1; 105/2

Query Match            19.1%; Score 307; DB 2; Length 262;  
Best Local Similarity 34.8%; Pred. No. 5,4e-18;  
Matches    80; Conservative    48; Mismatches    70; Indels    32; Gaps    8;

Dy         10 PLLLLGISGLLFITYOEVSR-----MKSAAVNQRVVITDAISGLGKECARVFHTGGAR 63  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         15 PALVVPIS-LTVAYVGLINRIIPGAHNLPKLDVKNKVVITGTASSGSLGKSLAFELYKGAQ 73  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Dy         64 LVLCGRMERLENLYDALISVADPSKT---TPKVLDLSDISCVP--DVAKEDVDC 116  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db         74 VILLARSTEKLEICEEL-----KETPLNQNEBIYYFPIITSEQA PMAEIR----- 122  
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Query 117 GCVADILNNMSVKKGPAHKSILELDKIMDANFGPFTLKALLPNMISRTGOIVLV 176  
Db 123 ---VDILNNMGMSNRKSGCQDTTMEIHKQAMETNIFGHVHTVQALLSKL--SPDGIIVT 177  
177 NNIOGKFGIPFTTYAASKHAALGFPDCLRAVEEYDVIVSTVSPFTIRS 226  
Db 178 SSIQKVALIPRGSYGASKHALOGYFDCLRAHKMLHLIV--VSAGYIWT 225

RESULT 4  
E83152  
probable short-chain dehydrogenase PA3957 [imported] - Pseudomonas aeruginosa (strain I  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: E83152  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; E  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lin  
.: Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: E83152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-278 <STO>  
A:Cross-references: UNIPROT:O9HK59; GB:AE004813; GB:AE004091; NID:G9950134; PIDN:AA072  
C:Genetics:  
C:Experimental source: strain PA01  
C:Superfamily: 11beta-hydroxysteroid dehydrogenase; short-chain alcohol dehydrogenase h

Query Match 18.6%; Score 299.5; DB 2; Length 278;  
Best Local Similarity 28.6%; Pred. No.2.5e-17;  
Matches 87; Conservative 57; Mismatches 99; Indels 61; Gaps 8;

Db 28 RLWKSNAVONRVVITDAISGLGKRCARVFTTGARVLVCGKNMERLENLY----- 78  
4 KVFSSQAVRHKVIVLVSGSGSIGRALRLFRAGARLALDLDDQALDSLVOHLRDLHGG 63  
79 DALISVADPSTKTFPKVLVLDLSDISGYPDVAKVLDGCGVDLILNNAASVVKPAHKK 138  
Db 64 EAL-----GRCVDADADAVERAVALAVERFGGIDVLVNNAGITHRGFEAT 110  
139 SLELDKIMDANFGPFTLKALLPNMISRTGOIVLVNNIOGKFGIPFTTYAASKHAA 198  
Db 111 GLGVFRKMAVNFQAVHCTPAALPSLERR--GQIVLVGSLTGFAPLRYRSVNNASKHAL 169  
199 LGFPDCLRAVEEYDVIVSTVSPFTIR-----SYHVPDQ--GNMEASIKWKF 244  
Db 170 HGLFDTLMELEGVSVTLACPGFTATDLRKNALVGDGSVTRQPVQVLGSCVAA----- 224  
245 FRKLTGYCHPEVAEVEWNRITVRKKQEVFMANPIPKAVVYRTFPEEPFAVAGVEX 304  
Db 225 -----PEVAEALFQGAARRRRLLVLISVNNWRARLLAR--FPPLF-----EK 265  
305 LNVP 308  
Db 266 LNVP 269

RESULT 5  
AG1983  
hypothetical protein ali1418 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AG1983  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaoka, S.; Watanabe, A.; Iriuguchi  
R.Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata,  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Takazawa Nitrogen-fixing Cyanobacterium Ar  
A:Reference number: AB1807; MUID:21595285; PMID:11759840

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: August 12, 2005, 16:51:19 ; Search time 43 Seconds  
(without alignments)  
539,904 Million cell updates/sec

Title: US-10-664-506A-5  
Perfect score: 1607  
Sequence: 1 MGVAMMLPPLLIGISGLL.....FFFAVACGVKLVNPEEG 311

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/prodata/1/1aa/5A COMB pep: \*  
2: /cgn2\_6/prodata/1/1aa/5B COMB pep: \*  
3: /cgn2\_6/prodata/1/1aa/6A COMB pep: \*  
4: /cgn2\_6/prodata/1/1aa/6B COMB pep: \*  
5: /cgn2\_6/prodata/1/1aa/6C COMB pep: \*  
6: /cgn2\_6/prodata/1/1aa/backfile1 pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1607	100.0	311	4	US-09-634-955B-5 <i>2404</i> Sequence 5, Appl1
2	1607	100.0	311	4	US-09-816-760-5 <i>2404</i> Sequence 5, Appl1
3	1607	100.0	311	4	US-09-838-561-5 <i>2404</i> Sequence 5, Appl1
4	1607	91.3	311	4	US-09-838-561-15 Sequence 15, Appl1
5	651.5	40.5	310	4	US-09-907-794A-153 Sequence 153, App
6	651.5	40.5	310	4	US-09-905-125A-153 Sequence 153, App
7	651.5	40.5	310	4	US-09-902-775A-153 Sequence 153, App
8	651.5	40.5	310	4	US-09-906-770-153 Sequence 153, App
9	651.5	40.5	310	4	US-09-903-603A-153 Sequence 153, App
10	651.5	40.5	310	4	US-09-904-920A-153 Sequence 153, App
11	651.5	40.5	310	4	US-09-909-064-153 Sequence 153, App
12	651.5	40.5	310	4	US-09-905-381A-153 Sequence 153, App
13	651.5	40.5	310	4	US-09-906-618-153 Sequence 153, App
14	651.5	40.5	325	4	US-09-311-021-74 Sequence 74, Appl1
15	651.5	40.5	325	4	US-09-806-536A-10 Sequence 10, Appl1
16	650.5	40.5	350	4	US-09-149-476-452 Sequence 452, App
17	300.5	18.7	271	4	US-09-328-352-4259 Sequence 4259, App
18	300	18.7	318	4	US-09-148-545-192 Sequence 192, App
19	295.5	18.4	339	4	US-09-148-545-248 Sequence 248, App
20	282.5	17.6	256	4	US-09-902-540-16103 Sequence 16103, A
21	282.5	17.6	275	4	US-09-902-540-9880 Sequence 9880, App
22	264	16.4	348	4	US-09-902-540-10937 Sequence 10937, A
23	256.5	16.0	292	4	US-09-252-991A-22228 Sequence 22228, A
24	252	15.7	295	3	US-09-026-482B-2 Sequence 2, Appl1
25	249.5	15.5	298	4	US-09-328-352-6138 Sequence 6138, App
26	240	14.9	252	4	US-09-134-000C-4224 Sequence 4224, App
27	237	14.7	360	4	US-09-949-016-10048 Sequence 10048, A

28	236.5	14.7	270	4	US-09-902-540-15606 Sequence 15606, A
29	233	14.5	332	4	US-09-949-016-6161 Sequence 6161, App
30	228.5	14.2	226	4	US-09-902-540-14378 Sequence 14378, A
31	227	14.1	253	4	US-09-583-110-3681 Sequence 3681, App
32	227	14.1	284	4	US-09-107-433-4311 Sequence 4311, App
33	224.5	14.0	257	4	US-09-489-039A-8162 Sequence 8162, App
34	224	13.9	238	4	US-08-586-664-1 Sequence 1, Appl1
35	224	13.9	248	3	US-09-385-028-11 Sequence 11, Appl1
36	224	13.9	248	4	US-09-726-614-11 Sequence 11, Appl1
37	224	13.9	248	4	US-09-385-040-11 Sequence 11, Appl1
38	218.5	13.6	249	4	US-09-134-000C-5219 Sequence 5219, App
39	218	13.6	272	4	US-09-107-433-4092 Sequence 4092, App
40	215	13.4	271	4	US-09-489-039A-12822 Sequence 12822, A
41	215	13.4	282	4	US-09-489-039A-12880 Sequence 12880, A
42	212	13.2	361	4	US-09-270-767-43794 Sequence 43794, A
43	209	13.0	231	3	US-09-134-001C-4397 Sequence 4397, App
44	209	13.0	300	4	US-09-907-794A-159 Sequence 159, App
45	209	13.0	300	4	US-09-866-028-37 Sequence 37, Appl1

## ALIGNMENTS

RESULT 1	US-09-634-955B-5	Sequence 5, Application US/09634955B
Patent No. 6511834		
GENERAL INFORMATION:		
APPLICANT: Meyers, Rachel		
TITLE OF INVENTION: 32142, 21486, 25964, 21686, NOVEL HUMAN DEHYDROGENASE		
FILE REFERENCE: NMT-134		
CURRENT APPLICATION NUMBER: US/09/634, 955B		
CURRENT FILING DATE: 2000-08-08		
PRIOR APPLICATION NUMBER: 60/192,002		
PRIOR FILING DATE: 2000-03-24		
NUMBER OF SEQ ID NOS: 35		
SOFTWARE: FASTSEQ for Windows Version 4.		
SEQ ID NO 5		
LENGTH: 311		
TYPE: PRT		
ORGANISM: Homo sapiens		
US-09-634-955B-5		
Query Match	100.0%	Score 1607; DB 4; Length 311;
Best Local Similarity	100.0%	Pred. No. 8.2e-171;
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	MGVAMMLPPLLIGISGLLFTIOEVSRLMSKSAVONKVVITDAISGLKECARVHTG 60
DB	1	MGVAMMLPPLLIGISGLLFTIOEVSRLMSKSAVONKVVITDAISGLKECARVHTG 60
QY	61	GARLVACGKWELELYDALISVADPSKTFPKVLVLLDSDISCVDAKEVLDGCV 120
DB	61	GARLVACGKWELELYDALISVADPSKTFPKVLVLLDSDISCVDAKEVLDGCV 120
QY	121	DILINNASYKVGPAKISLELDKIMDANYGPIITLTKALLPNNISRTGQIVLVNIIQ 180
DB	121	DILINNASYKVGPAKISLELDKIMDANYGPIITLTKALLPNNISRTGQIVLVNIIQ 180
QY	181	GKRGIPRTTYAASKAALGFDDCLAAVEEDVISTSPFIRIRHYHYPEGKNEAST 240
DB	181	GKRGIPRTTYAASKAALGFDDCLAAVEEDVISTSPFIRIRHYHYPEGKNEAST 240
QY	241	WKEPFKLTGVHPVAAEVRWTRRKQOEVNANPIKAAVYVATPPEFFFAVACG 300
DB	241	WKEPFKLTGVHPVAAEVRWTRRKQOEVNANPIKAAVYVATPPEFFFAVACG 300
QY	301	VKEKLVNPEEG 311
DB	301	VKEKLVNPEEG 311

RESULT 2  
US-09-816-760-5  
Sequence 5, Application US/09816760  
Patent No. 661355  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
APPLICANT: Cook, William James  
APPLICANT: Williamson, Mark  
APPLICANT: Rudolph-Owen, Laura A.  
TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL HUMAN DEHYDROGENASE  
TITLE OF INVENTION: MOLECULES AND USES THEREFOR  
FILE REFERENCE: MNI-134CP  
CURRENT APPLICATION NUMBER: US/09/816,760  
CURRENT FILING DATE: 2001-03-23  
PRIOR APPLICATION NUMBER: 09/634,955  
PRIOR FILING DATE: 2000-08-08  
PRIOR APPLICATION NUMBER: 60/192,002  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 311  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-816-760-5

Query Match 100.0%; Score 1607; DB 4; Length 311;

Best Local Similarity 100.0%; Pred. No. 8,2e-171; Mismatches 0; Indels 0; Gaps 0;

Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGWAMMLPLLLIGSLFTYQEVSRMSKSAVQKVVITDAISGLGKRCARVPHFG 60  
DB 1 MGWAMMLPLLLIGSLFTYQEVSRMSKSAVQKVVITDAISGLGKRCARVPHFG 60  
QY 61 GARVLGCGKMERLENYDALISVADPSKTFPTPKVLDDISICVDPVAKEVLDYCGV 120  
DB 61 GARVLGCGKMERLENYDALISVADPSKTFPTPKVLDDISICVDPVAKEVLDYCGV 120  
QY 121 DILINNASVKYKGPAAKISLELDKIMDANYFGPITLTKALLPMMISRRTQIVLVNNIQ 180  
DB 121 DILINNASVKYKGPAAKISLELDKIMDANYFGPITLTKALLPMMISRRTQIVLVNNIQ 180  
QY 181 GKFGIPRTTYAASKHAGLGFDCRAVEEYDVISTVSPFTIRSYHYVPEQGNWEASI 240  
DB 181 GKFGIPRTTYAASKHAGLGFDCRAVEEYDVISTVSPFTIRSYHYVPEQGNWEASI 240  
QY 241 WKFFPKLTGYVHPVEVAEVMRTVRKKQEVFMANIPKAAVYVTRTFPEFFAVVACG 300  
DB 241 WKFFPKLTGYVHPVEVAEVMRTVRKKQEVFMANIPKAAVYVTRTFPEFFAVVACG 300  
QY 301 VKEKLVPEEG 311  
DB 301 VKEKLVPEEG 311

RESULT 3  
US-09-838-561-5  
Sequence 5, Application US/09838561  
Patent No. 6627423  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
APPLICANT: Cook, William James  
APPLICANT: Williamson, Mark  
APPLICANT: Rudolph-Owen, Laura A.  
TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL DEHYDROGENASE  
TITLE OF INVENTION: MOLECULES AND USES THEREFOR  
FILE REFERENCE: MNI-134CP2  
CURRENT APPLICATION NUMBER: US/09/838,561  
CURRENT FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 09/816,760  
PRIOR FILING DATE: 2001-03-23

PRIOR APPLICATION NUMBER: 09/634,955  
PRIOR FILING DATE: 2000-08-08  
PRIOR APPLICATION NUMBER: 60/192,002  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 311  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-838-561-5

Query Match 100.0%; Score 1607; DB 4; Length 311;  
Best Local Similarity 100.0%; Pred. No. 8,2e-171; Mismatches 0; Indels 0; Gaps 0;

Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGWAMMLPLLLIGSLFTYQEVSRMSKSAVQKVVITDAISGLGKRCARVPHFG 60  
DB 1 MGWAMMLPLLLIGSLFTYQEVSRMSKSAVQKVVITDAISGLGKRCARVPHFG 60  
QY 61 GARVLGCGKMERLENYDALISVADPSKTFPTPKVLDDISICVDPVAKEVLDYCGV 120  
DB 61 GARVLGCGKMERLENYDALISVADPSKTFPTPKVLDDISICVDPVAKEVLDYCGV 120  
QY 121 DILINNASVKYKGPAAKISLELDKIMDANYFGPITLTKALLPMMISRRTQIVLVNNIQ 180  
DB 121 DILINNASVKYKGPAAKISLELDKIMDANYFGPITLTKALLPMMISRRTQIVLVNNIQ 180  
QY 181 GKFGIPRTTYAASKHAGLGFDCRAVEEYDVISTVSPFTIRSYHYVPEQGNWEASI 240  
DB 181 GKFGIPRTTYAASKHAGLGFDCRAVEEYDVISTVSPFTIRSYHYVPEQGNWEASI 240  
QY 241 WKFFPKLTGYVHPVEVAEVMRTVRKKQEVFMANIPKAAVYVTRTFPEFFAVVACG 300  
DB 241 WKFFPKLTGYVHPVEVAEVMRTVRKKQEVFMANIPKAAVYVTRTFPEFFAVVACG 300  
QY 301 VKEKLVPEEG 311  
DB 301 VKEKLVPEEG 311

RESULT 4  
US-09-838-561-15  
Sequence 15, Application US/09838561  
Patent No. 6627423  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
APPLICANT: Cook, William James  
APPLICANT: Williamson, Mark  
APPLICANT: Rudolph-Owen, Laura A.  
TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL DEHYDROGENASE  
TITLE OF INVENTION: MOLECULES AND USES THEREFOR  
FILE REFERENCE: MNI-134CP2  
CURRENT APPLICATION NUMBER: US/09/838,561  
CURRENT FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 09/816,760  
PRIOR FILING DATE: 2001-03-23  
PRIOR APPLICATION NUMBER: 09/634,955  
PRIOR FILING DATE: 2000-08-08  
PRIOR APPLICATION NUMBER: 60/192,002  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 311  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-838-561-15

Query Match 91.3%; Score 1467; DB 4; Length 311;  
Best Local Similarity 91.3%; Pred. No. 3,6e-155; Mismatches 284; Conservative 8; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2005, 16:41:03 ; Search time 165 Seconds  
(without alignments)  
728.985 Million cell updates/sec

Title: US-10-664-506A-5  
Perfect score: 1607  
Sequence: 1 MGWAMLMPLLLIGISGL.....FPFAVAVGVKELNPEEG 311

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1607	100.0	311	4 AAB47593	AAB47593 Human DHD
2	1607	100.0	311	6 ABP58046	ABP58046 Human deh
3	1607	100.0	311	8 ADG40199	ADG40199 Human deh
4	1601	99.6	311	3 AAB08911	AAB08911 Human sec
5	1601	99.6	345	3 AAB08948	AAB08948 Human sec
6	1467	91.3	311	6 ABP58047	ABP58047 Mouse deh
7	1467	91.3	311	8 ADG40201	ADG40201 Mouse deh
8	1258	78.3	246	8 ADM33425	ADM33425 Human PRO
9	651.5	40.5	310	3 AAB27650	AAB27650 Human PRO
10	651.5	40.5	310	3 ADC78473	ADC78473 Human PRO
11	651.5	40.5	310	4 AAB80238	AAB80238 Human PRO
12	651.5	40.5	310	5 ABB84835	ABB84835 Human PRO
13	651.5	40.5	310	5 ABB95441	ABB95441 Human ang
14	651.5	40.5	310	6 ABU71616	ABU71616 Human PRO
15	651.5	40.5	310	6 ABU71471	ABU71471 Human PRO
16	651.5	40.5	310	6 ABU71917	ABU71917 Human sec
17	651.5	40.5	310	6 ABO01800	ABO01800 Novel hum
18	651.5	40.5	310	6 ABO43373	ABO43373 Human sec
19	651.5	40.5	310	6 ABO47388	ABO47388 Human sec
20	651.5	40.5	310	6 ABU64525	ABU64525 Human sec
21	651.5	40.5	310	6 ABU67371	ABU67371 Human sec
22	651.5	40.5	310	6 ABO14891	ABO14891 Human sec
23	651.5	40.5	310	6 ABU69648	ABU69648 Novel hum
24	651.5	40.5	310	6 ABO14830	ABO14830 Human sec
25	651.5	40.5	310	6 ADB29358	ADB29358 Human sec

26	651.5	40.5	310	6 ADA18214	ADA18214 Human sec
27	651.5	40.5	310	6 ABO32782	ABO32782 Human sec
28	651.5	40.5	310	6 ABO34842	ABO34842 Human PRO
29	651.5	40.5	310	6 ADA16189	ADA16189 Human sec
30	651.5	40.5	310	6 ADA42334	ADA42334 Human sec
31	651.5	40.5	310	6 ABO17520	ABO17520 Human PRO
32	651.5	40.5	310	7 ADA16613	ADA16613 Human sec
33	651.5	40.5	310	7 ADA13042	ADA13042 Human sec
34	651.5	40.5	310	7 ADA1910	ADA1910 Human sec
35	651.5	40.5	310	7 ADA17257	ADA17257 Human sec
36	651.5	40.5	310	7 ADA42760	ADA42760 Human sec
37	651.5	40.5	310	7 ABO17581	ABO17581 Human PRO
38	651.5	40.5	310	7 ADB77679	ADB77679 Human sec
39	651.5	40.5	310	7 ADB74815	ADB74815 Human sec
40	651.5	40.5	310	7 ADC28461	ADC28461 Human sec
41	651.5	40.5	310	7 ADC39661	ADC39661 Human sec
42	651.5	40.5	310	7 ADC40175	ADC40175 Human sec
43	651.5	40.5	310	7 ADC19003	ADC19003 Human sec
44	651.5	40.5	310	7 ADC34229	ADC34229 Human sec
45	651.5	40.5	310	7 ADC29354	ADC29354 Human sec

## ALIGNMENTS

RESULT 1	
AAB47593	
ID AAB47593 standard; protein; 311 AA.	
AC AAB47593;	
XX	
DT 07-JAN-2002 (first entry)	
XX	
DE Human DHDR-2.	
XX	
KW Human; dehydrogenase; DHDR; cellular proliferation; CNS; cardiac;	
KW muscular; cellular growth; differentiation; migration; neurological;	
KW immune system; ss.	
XX	
OS Homo sapiens.	
XX	
PH Key	Location/Qualifiers
FT Peptide	1..18
FT	/label= Signal peptide
FT Domain	7..23
FT	/label= Transmembrane domain
FT Domain	38..227
FT	/label= Short chain dehydrogenase domain
FT Domain	99..219
FT	/label= Oxidoreductase protein dehydrogenase domain
XX	
PN W0200172976-A2.	
XX	
PD 04-OCT-2001.	
XX	
PF 23-MAR-2001; 2001WO-US009613.	
XX	
PR 24-MAR-2000; 2000US-0192002P.	
PR 08-AUG-2000; 2000US-00634955.	
XX	
PA (MIL-) MILLENNIUM PHARM INC.	
XX	
PI Meyers R, Cook WJ, Williamson M, Rudolph-Owen LA;	
XX	
DR WPI; 2001-639228/73.	
DR N-PSDB; AAA43561, AAA43565.	
XX	
PT New human dehydrogenase molecules designated DHDR-1, DHDR-2, DHDR-3 and	
PT DHDR-4 for finding treatment for dehydrogenase-associated disorders such	
PT as viral and cellular proliferation disorders.	
XX	
PS Claim 2, Fig 6; 83pp, English.	
XX	

CC The sequences given in AAH43560-67 encode human dehydrogenase (DHDR)-1-4.  
 CC Modulators of DHDR activity are used to treat a viral or cellular  
 CC proliferation disorder. The DHDR proteins can also be used to treat  
 CC disorders characterized by over or under production of DHDR proteins, for  
 CC example dehydrogenase-associated disorders such as CNS, cardiac,  
 CC muscular, cellular growth, differentiation or migration, neurological,  
 CC immune, humoral or viral disorders. The molecules can be used in  
 CC pharmacogenetics  
 CC  
 CC Sequence 311 AA;  
 CC  
 CC

Query Match 100.0%; Score 1607; DB 4; Length 311;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-163;  
 Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGWAMMLPPLLLIGSLFTYQEVSRMSKSAVONKVVITDAISGLGECARVHTG 60  
 DB 1 MGWAMMLPPLLLIGSLFTYQEVSRMSKSAVONKVVITDAISGLGECARVHTG 60  
 QY 61 GARLVLCGKMERLENYDALISVADPSKTFPKVLVLDLSDISCPDVAKEVLDYGCY 120  
 DB 61 GARLVLCGKMERLENYDALISVADPSKTFPKVLVLDLSDISCPDVAKEVLDYGCY 120  
 QY 121 DILINNASVKVKGPAHKISLELDKIMDANYFGPITLTKALLPMISRTGOIVLVNNIQ 180  
 DB 121 DILINNASVKVKGPAHKISLELDKIMDANYFGPITLTKALLPMISRTGOIVLVNNIQ 180  
 QY 181 GKFGIPRTTYAASKRAAGFPDCLRAVEEYDVISTVSPTFRSHVHYVEQGNWENSI 240  
 DB 181 GKFGIPRTTYAASKRAAGFPDCLRAVEEYDVISTVSPTFRSHVHYVEQGNWENSI 240  
 QY 241 WKFFPRKLTGYGVHPEVAEEMRTVRRKKQEVFMANPIPKAAVYVTRFFPEFFAVVACG 300  
 DB 241 WKFFPRKLTGYGVHPEVAEEMRTVRRKKQEVFMANPIPKAAVYVTRFFPEFFAVVACG 300  
 QY 301 VKEKLVNPEEG 311  
 DB 301 VKEKLVNPEEG 311

RESULT 2  
 ABB58046  
 ID ABB58046 standard; protein; 311 AA.  
 XX

AC ABB58046;  
 XX  
 XX 03-MAR-2003 (first entry)  
 XX

DE Human dehydrogenase-2 (DHDR-2).

XX Dehydrogenase-2; DHDR-2; human; thermogenesis; body weight disorder;  
 KM obesity; overweight; anorexia; cachexia; insulin resistance; diabetes;  
 KM antidiabetic; anorectic; immunomodulator; antidepressant; gene therapy;  
 KM enzyme.

XX Homo sapiens.

OS MO200290576-A1.

PN 14-NOV-2002.

PD 09-MAY-2002; 2002WO-US014882.

PF 09-MAY-2001; 2001US-0289917P.

PR (MILL-) MILLENNIUM PHARM INC.

PA Gímeno R, Spiegelman BM;

PT WPI; 2003-103521/09.

DR N-PSDB; ABV76050.

XX Identifying a compound useful for treating a bodyweight disorder, e.g.

PT obesity or diabetes, comprises assaying the ability of the compound to  
 PT modulate dehydrogenase-2 (DHDR-2) nucleic acid expression or DHDR-2  
 PT polypeptide activity.

Claim 13; Fig 3; 73p; English.

CC The present sequence is the protein sequence of human dehydrogenase-2  
 CC (DHDR-2). The invention is based on the discovery that expression of the  
 CC DHDR-2 gene is downregulated during cold exposure, i.e. during  
 CC thermogenesis, and is also downregulated in cells expressing thermogenic  
 CC coactivator PGC-1. DHDR-2 molecules, as part of the thermogenic  
 CC signalling pathway, modulate thermogenesis and are useful as targets and  
 CC therapeutic agents for the modulation of thermogenesis, e.g. expenditure  
 CC of energy and the treatment of body weight disorders. A claimed method  
 CC for identifying a compound capable of treating a body weight disorder  
 CC comprises assaying the ability of the compound to modulate DHDR-2 nucleic  
 CC acid expression or DHDR-2 polypeptide activity, e.g. by detecting  
 CC mitochondrial activity of a cell or detecting thermogenesis in a cell.  
 CC The body weight disorder is obesity, overweight, diabetes, insulin  
 CC resistance, cachexia or anorexia. A claimed method for identifying a  
 CC compound capable of modulating thermogenesis involves contacting a cell  
 CC which expresses DHDR-2, especially a primary muscle cell, a C2C12 myocyte  
 CC and a C2C12 myotube, with a test compound and assaying the ability of the  
 CC compound to modulate DHDR-2 expression or activity. The modulator is a  
 CC DHDR-2 polypeptide, its fragment or variant, an antisense DHDR-2 nucleic  
 CC acid, a ribozyme, or a DHDR-2 nucleic acid. The DHDR-2 modulator is used  
 CC in a claimed method of treating a body weight disorder, and may be  
 CC administered using a gene therapy vector. The methods and DHDR-2  
 CC modulators are useful for diagnosing, preventing and treating a body  
 CC weight disorder or for modulating thermogenesis  
 CC  
 CC Sequence 311 AA;  
 CC  
 CC

Query Match 100.0%; Score 1607; DB 6; Length 311;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-163;  
 Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGWAMMLPPLLLIGSLFTYQEVSRMSKSAVONKVVITDAISGLGECARVHTG 60  
 DB 1 MGWAMMLPPLLLIGSLFTYQEVSRMSKSAVONKVVITDAISGLGECARVHTG 60  
 QY 61 GARLVLCGKMERLENYDALISVADPSKTFPKVLVLDLSDISCPDVAKEVLDYGCY 120  
 DB 61 GARLVLCGKMERLENYDALISVADPSKTFPKVLVLDLSDISCPDVAKEVLDYGCY 120  
 QY 121 DILINNASVKVKGPAHKISLELDKIMDANYFGPITLTKALLPMISRTGOIVLVNNIQ 180  
 DB 121 DILINNASVKVKGPAHKISLELDKIMDANYFGPITLTKALLPMISRTGOIVLVNNIQ 180  
 QY 181 GKFGIPRTTYAASKRAAGFPDCLRAVEEYDVISTVSPTFRSHVHYVEQGNWENSI 240  
 DB 181 GKFGIPRTTYAASKRAAGFPDCLRAVEEYDVISTVSPTFRSHVHYVEQGNWENSI 240  
 QY 241 WKFFPRKLTGYGVHPEVAEEMRTVRRKKQEVFMANPIPKAAVYVTRFFPEFFAVVACG 300  
 DB 241 WKFFPRKLTGYGVHPEVAEEMRTVRRKKQEVFMANPIPKAAVYVTRFFPEFFAVVACG 300  
 QY 301 VKEKLVNPEEG 311  
 DB 301 VKEKLVNPEEG 311

RESULT 3  
 ADG40199

ID ADG40199 standard; protein; 311 AA.

AC ADG40199;

XX 26-FEB-2004 (first entry)

XX Human dehydrogenase-2 (DHDR-2).

XX Thermogenesis; dehydrogenase-2 modulator; DHDR-2; body weight disorder;



-continued

aac aac atc caa gcg aag ttt gga atc ccg ttc cgc aca gct tat gca	576
Asn Asn Ile Gln Ala Lys Phe Gly Ile Pro Phe Arg Thr Ala Tyr Ala	
180 185 190	
gcc tct aag cat gcc gtc atg ggc ttc ttt gac tgc ctc cga gcc gag	624
Ala Ser Lys His Ala Val Met Gly Phe Phe Asp Cys Leu Arg Ala Glu	
195 200 205	
gtt gag gaa tac gat gtt gtg gtc agc acc gtg agc cca act ttc atc	672
Val Glu Glu Tyr Asp Val Val Val Ser Thr Val Ser Pro Thr Phe Ile	
210 215 220	
cgc tcc tac cgt gct tcc cct gag caa aga aac tgg gag aca tcc att	720
Arg Ser Tyr Arg Ala Ser Pro Glu Gln Arg Asn Trp Glu Thr Ser Ile	
225 230 235 240	
tgt aaa ttc ttc tgc agg aag cta gcc tat ggc gtg cac ccg gtg gag	768
Cys Lys Phe Phe Cys Arg Lys Leu Ala Tyr Gly Val His Pro Val Glu	
245 250 255	
gtg gct gag gaa gtg atg cgc aca gta cgg agg aag aag caa gag gtg	816
Val Ala Glu Glu Val Met Arg Thr Val Arg Arg Lys Lys Gln Glu Val	
260 265 270	
ttc atg gcc aac ccg gtt cct aag gct gcc gtg ttc atc cgc acc ttc	864
Phe Met Ala Asn Pro Val Pro Lys Ala Ala Val Phe Ile Arg Thr Phe	
275 280 285	
ttc cct gag ttc ttc ttc gct gtg gtg gcc tgt ggg gtg aag gag aag	912
Phe Pro Glu Phe Phe Phe Ala Val Val Ala Cys Gly Val Lys Glu Lys	
290 295 300	
ctc aat gtc cca gaa gag ggt	933
Leu Asn Val Pro Glu Glu Gly	
305 310	

Att + Dist

Refined

What is claimed:

1. An isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:4, or a full complement thereof.

2. An isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:6, or a full complement thereof.

3. An isolated nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:4, or a full complement thereof.

4. An isolated nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:6, or a full complement thereof.

5. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:5, or a full complement thereof.

6. An isolated nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:5, or a full complement thereof.

7. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number PTA-1845, or a full complement thereof.

8. An isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:4 or SEQ ID NO:6, or a full complement thereof, wherein said nucleic acid molecule encodes a polypeptide having a dehydrogenase activity.

9. An isolated nucleic acid molecule consisting of a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:4 or SEQ ID NO:6, or a full complement thereof, wherein said nucleic acid molecule encodes a polypeptide having a dehydrogenase activity.

10. An isolated nucleic acid molecule encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:5, or a full complement thereof, wherein said polypeptide has a dehydrogenase activity.

11. An isolated nucleic acid molecule encoding a polypeptide consisting of an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NO:5, or a full complement thereof, wherein said polypeptide has a dehydrogenase activity.

12. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1-4, and a nucleotide sequence encoding a heterologous polypeptide.

13. A vector comprising the nucleic acid molecule of any one of claims 1-4.

14. The vector of claim 13, which is an expression vector.

15. A host cell transfected with the expression vector of claim 14.

16. A method of producing a polypeptide having dehydrogenase activity comprising culturing the host cell of claim 15 under conditions in which the nucleic acid molecule is expressed, thereby expressing the polypeptide.

\* \* \* \* \*

-continued

1	5	10	15
Gln Asn Gln Val Ala Val Val Thr Gly Gly Ala Thr Gly Ile Gly Lys	20	25	30
Ala Ile Ser Arg Glu Leu Leu His Leu Gly Cys Asn Val Val Ile Ala	35	40	45
Ser Arg Lys Leu Asp Arg Leu Thr Ala Ala Val Asp Glu Leu Arg Ala	50	55	60
Ser Gln Pro Pro Ser Ser Ser Thr Gln Val Thr Ala Ile Gln Cys Asn	65	70	75
Ile Arg Lys Glu Glu Glu Val Asn Asn Leu Val Lys Ser Thr Leu Ala	85	90	95
Lys Tyr Gly Lys Ile Asn Phe Leu Val Asn Asn Ala Gly Gly Gln Phe	100	105	110
Met Ala Pro Ala Glu Asp Ile Thr Ala Lys Gly Trp Gln Ala Val Ile	115	120	125
Glu Thr Asn Leu Thr Gly Thr Phe Tyr Met Cys Lys Ala Val Tyr Asn	130	135	140
Ser Trp Met Lys Asp His Gly Gly Ser Ile Val Asn Ile Ile Val Leu	145	150	155
Leu Asn Asn Gly Phe Pro Thr Ala Ala His Ser Gly Ala Ala Arg Ala	165	170	175
Gly Val Tyr Asn Leu Thr Lys Thr Met Ala Leu Thr Trp Ala Ser Ser	180	185	190
Gly Val Arg Ile Asn Cys Val Ala Pro Gly Thr Ile Tyr Ser Gln Thr	195	200	205
Ala Val Asp Asn Tyr Gly Glu Leu Gly Gln Thr Met Phe Glu Met Ala	210	215	220
Phe Glu Asn Ile Pro Ala Lys Arg Val Gly Leu Pro Glu Glu Ile Ser	225	230	235
Pro Leu Val Cys Phe Leu Leu Ser Pro Ala Ala Ser Phe Ile Thr Gly	245	250	255
Gln Leu Ile Asn Val Asp Gly Gly Gln Ala Leu Tyr Thr Arg Asn Phe	260	265	270
Thr Ile Pro Asp His Asp Asn Trp Pro Val Gly Ala Gly Asp Ser Ser	275	280	285
Phe Ile Lys Lys Val Lys Glu Ser Leu Lys Lys Gln Ala Arg Leu	290	295	300

What is claimed:

1. A method for identifying a compound which binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:2;

the method comprising:

- i) contacting the polypeptide, or a cell expressing the polypeptide with a test compound under conditions suitable for binding; and
- ii) detecting binding of the test compound to the polypeptide.

2. A method for identifying a compound which binds to a polypeptide consisting of the amino acid sequence of SEQ ID NO:2;

the method comprising:

- a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound under conditions suitable for binding; and
- b) detecting binding of the test compound to the polypeptide.

3. A method for identifying a compound which binds to a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:10 or 12, and wherein said polypeptide has a dehydrogenase activity;

the method comprising:

- i) contacting the polypeptide, or a cell expressing the polypeptide with a test compound under conditions suitable for binding; and
- ii) detecting binding of the test compound to the polypeptide.

4. A method for identifying a compound which binds to a polypeptide comprising an amino acid sequence which is at least 95% identical to the entire length of the amino acid sequence of SEQ ID NO:11, and wherein said polypeptide has a dehydrogenase activity;

119

the method comprising:

i) contacting the polypeptide, or a cell expressing the polypeptide with a test compound under conditions suitable for binding; and

ii) detecting binding of the test compound to the polypeptide.

5. A method for identifying a compound which binds to a polypeptide encoded by the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number PTA-3216;

the method comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound under conditions suitable for binding; and

b) detecting binding of the test compound to the polypeptide.

6. The method of any one of claims 1, 2, 3, 4, and 5, wherein said detection is by direct binding.

7. The method of claim 6, wherein said direct binding is determined by an immunoprecipitation.

120

8. The method of claim 6, wherein said direct binding is determined by a yeast two-hybrid assay.

9. The method of any one of claims 1, 2, 3, 4, or 5, wherein said detection is by the use of a competition binding assay.

10. The method of any one of claims 1, 2, 3, 4, or 5, wherein said cell is a tumor cell.

11. The method of any one of claims 1, 2, 3, 4, or 5, wherein said detection is by the use of an assay for an activity of the polypeptide consisting of the amino acid sequence of SEQ ID NO:11.

12. The method of claim 11, wherein said activity is cellular growth or proliferation.

13. The method of claim 11, wherein said activity is cellular signaling.

14. The method of claim 11, wherein said activity is modulation of viral gene expression.

\* \* \* \* \*

-continued

35 40 45

Ser Leu  
50

<210> SEQ ID NO 34  
<211> LENGTH: 57  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: hypothetical protein domain

<400> SEQUENCE: 34

Phe Tyr Lys Pro Asn Leu Glu Gln Tyr Gln His Arg Trp Thr Val Val  
1 5 10 15  
Ser Gly Gly Thr Asp Gly Ile Gly Lys Ala Tyr Thr Leu Glu Leu Ala  
20 25 30  
Lys Arg Gly Leu Arg Lys Phe Val Leu Ile Gly Arg Asn Pro Lys Lys  
35 40 45  
Leu Asp Ser Val Lys Ser Glu Ile Glu  
50 55

<210> SEQ ID NO 35  
<211> LENGTH: 45  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: glucose-1-dehydrogenase domain

<400> SEQUENCE: 35

Thr Leu Glu Met Ile Pro Ala Lys Glu Ile Gly Phe Ala Asp Gln Val  
1 5 10 15  
Ala Asn Val Ala Arg Phe Leu Cys Ser Asp Leu Ala Asp Tyr Ile His  
20 25 30  
Gly Thr Thr Ile Tyr Val Asp Gly Gly Met Thr Asn Tyr  
35 40 45

## What is claimed:

1. An isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:10, or a complement thereof.

2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:11, or a complement thereof.

3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC as Accession Number PTA-3216.

4. An isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:12, or a complement thereof.

5. An isolated nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:10, or a complement thereof.

6. An isolated nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:12, or a complement thereof.

7. An isolated nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:11, or a complement thereof.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, 5, 6, or 7, and a nucleotide sequence encoding a heterologous polypeptide.

9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, 5, 6, or 7.

10. The vector of claim 9, which is an expression vector.

11. A host cell transfected with the expression vector of claim 10.

12. A method of producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising or consisting of the amino acid sequence set forth in SEQ ID NO:11;

b) a polypeptide encoded by a nucleic acid molecule comprising or consisting of the nucleotide sequence set forth in SEQ ID NO:10 or SEQ ID NO:12; and

c) a polypeptide encoded by a nucleic acid molecule comprising or consisting of the nucleotide sequence contained in the plasmid deposited with ATCC as Accession Number PTA-3216,

the method comprising culturing the host cell of claim 10 in an appropriate culture medium to, thereby, produce the polypeptide.

13. A kit comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, 5, 6, or 7 and instructions for use.

\* \* \* \* \*